BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 7: e110459
doi: 10.3897/biss.7.110459

OPEN ACCESS

Conference Abstract

E-taxonomy, Open Science, and FAIR Principles: A comparison of how taxon descriptions are computerized

Adeline Kerner‡, Carla Levy‡

‡ UMR 7207 – Centre de Recherche en Paléontologie - Paris CNRS – Sorbonne Université - MNHN, Paris, France

Corresponding author: Adeline Kerner (kerner@mnhn.fr)

Received: 02 Aug 2023 | Published: 07 Aug 2023

Citation: Kerner A, Levy C (2023) E-taxonomy, Open Science, and FAIR Principles: A comparison of how taxon descriptions are computerized. Biodiversity Information Science and Standards 7: e110459. https://doi.org/10.3897/biss.7.110459

Abstract

Describing and naming are the first steps in a systematist's work in any biodiversity study. Comparing a single specimen with all the published descriptions is often a headache. The oldest solution to facilitate taxon identification was the creation of single-access identification keys. The first example of this is attributed to Lamarck (Lamarck et al. 1815). The first keys were, therefore, paper documents, mostly supplied with the monographs of the taxonomic groups concerned. Their creation and revision took a lot of time.

The emergence of computers quickly became a means of facilitating the sorting and accessing of descriptive data. Since the 1960s, new key forms have emerged, as well as programs to assist identification (Goodall 1968, Pankhurst and Aitchison 1975).

Descriptions need to be formalized and structured to be understandable by a computer. Thoughts on the structuring of descriptive data quickly emerged, working towards creating a standard. To address this, the Description Language for Taxonomy (DELTA) was ratified in 1986, followed by the SDD (Structured Descriptive Data) in 2005. The DELTA system relied on a data format closely tied to specific software programs. Any changes to the DELTA standard required corresponding modifications to a large amount of software code, making updates difficult. Alternative software packages using different data formats were developed but were not fully compatible with DELTA or with each other, preventing lossless

© Kerner A, Levy C. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

data translation. To overcome these challenges, the SDD data standard was created as an implementation-independent format through international collaboration, allowing for extensibility without disrupting applications and ensuring continued support even if specific implementing programs were discontinued.

Since the adoption of the SDD, numerous software programs and other projects have been developed to provide identification methods (e.g., XPER3, Lucid, AbaTax). Unfortunately, the SDD is not always used, as it may not fit in with the unique objectives of each of them.

New tools with other structures and methods are emerging, which is great. However, in the context of FAIR data (Findable, Accessible, Interoperable, and Reusable digital assets), no longer using a standard seems to go against the flow. Have our needs changed? How to maintain models for sharing descriptive data that are structured enough to become knowledge allowing comparison, identification, data-mining, and analysis?

To provide a starting point for thinking about the future of SDD, we present a comparison of several identification tools. The comparison focuses on their use of descriptors, items/taxa, identifiers, media, logical operators, their expression of doubt, polymorphism, and "multi-states," but also the files supported and the functionalities from the user's point of view.

Keywords

identification key, SDD structured descriptive data

Presenting author

Adeline Kerner

Presented at

TDWG 2023

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Goodall D (1968) Identification by Computer. BioScience 18 (6): 485-488. https://doi.org/10.2307/1294274
- Lamarck J, Candolle APd, Dezauche AG, Sève JEd, Sève JEd, Desray J (1815) Flore française, ou, Descriptions succinctes de toutes les plantes qui croissent naturellement en France : disposées selon une nouvelle méthode d'analyse, et précédées par un

exposé des principes élémentaires de la botanique. Chez Desray https://doi.org/10.5962/bhl.title.112968

- Pankhurst RJ, Aitchison RR (1975) An on-line identification program. In: Pankhurst RJ (Ed.) Biological identification with computers. Academic Press